CLAIMS

What is claimed is:

1. A machine-implemented method comprising:

extracting portions from segment boundary regions of a plurality of speech segments, each segment boundary region based on a corresponding initial unit boundary;

creating feature vectors that represent the portions in a vector space;

for each of a plurality of potential unit boundaries within each segment boundary region, determining an average discontinuity based on distances between the feature vectors; and

for each segment, selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

2. The machine-implemented method of claim 1, further comprising:

if all of the new unit boundaries are the same as the corresponding initial unit boundaries, setting the new unit boundaries as final unit boundaries for the segments.

3. The machine-implemented method of claim 1, further comprising:

if any of the new unit boundaries are different from the corresponding initial unit boundaries, iteratively:

setting the new unit boundary as the initial unit boundary, and

performing the extracting, the creating, the determining and the selecting, until all of the new unit boundaries are the same as the corresponding initial unit boundaries.

4. The machine-implemented method of claim 1, wherein the average discontinuity is determined over a plurality of concatenations.

5. The machine-implemented method of claim 1, wherein the initial unit boundary is in the middle of a phoneme.

6. The machine-implemented method of claim 1, wherein each potential unit boundary defines two candidate units for each speech segment.

7. The machine-implemented method of claim 6, wherein a concatenation of the plurality of concatenations includes a candidate unit of a first segment linked to a candidate unit of a second segment.

8. The machine-implemented method of claim 6, wherein the plurality of concatenations includes all combinations of a first candidate unit of each segment with a second candidate unit of each segment.

9. The machine-implemented method of claim 1, wherein the plurality of speech segments includes speech segments which end in the middle of a first phoneme, and speech segments which begin in the middle of a first phoneme.

10. The machine-implemented method of claim 9, wherein the plurality of speech segments are stored in a voice table.

11. The machine-implemented method of claim 1, further comprising:

recording speech input; and

identifying the speech segments within the speech input.

12. The machine-implemented method of claim 1, wherein the portions include centered pitch periods, the centered pitch periods derived from pitch periods of the segments.

13. The machine-implemented method of claim 12, wherein the feature vectors incorporate phase information of the portions.

14. The machine-implemented method of claim 13, wherein creating feature vectors comprises:

constructing a matrix W from the portions; and

decomposing the matrix W.

15. The machine-implemented method of claim 14, wherein the matrix W is a $(2(K-1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the potential unit boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R << (2(K\text{-}1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

16. The machine-implemented method of claim 15, wherein the centered pitch periods are symmetrically zero padded to N samples.

17. The machine-implemented method of claim 15, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

18. The machine-implemented method of claim 17, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

19. The machine-implemented method of claim 18, wherein a discontinuity $d(S_1,S_2)$ between two candidate units, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

20. The machine-implemented method of claim 19, wherein the same closeness measure, C, is used for optimizing unit boundaries and for unit selection.

21. A machine-readable medium having instructions to cause a machine to perform a machine-implemented method comprising:

extracting portions from segment boundary regions of a plurality of speech segments, each segment boundary region based on a corresponding initial unit boundary;

creating feature vectors that represent the portions in a vector space;

for each of a plurality of potential unit boundaries within each segment boundary region, determining an average discontinuity based on distances between the feature vectors; and

for each segment, selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

22. The machine-readable medium of claim 21, wherein the method further comprises:

if all of the new unit boundaries are the same as the corresponding initial unit boundaries, setting the new unit boundaries as final unit boundaries for the segments.

23. The machine-readable medium of claim 21, wherein the method further comprises:

if any of the new unit boundaries are different from the corresponding initial unit boundaries, iteratively:

setting the new unit boundary as the initial unit boundary, and

performing the extracting, the creating, the determining and the selecting, until all of the new unit boundaries are the same as the corresponding initial unit boundaries.

24. The machine-readable medium of claim 21, wherein the average discontinuity is determined over a plurality of concatenations.

25. The machine-readable medium of claim 21, wherein the initial unit boundary is in the middle of a phoneme.

26. The machine-readable medium of claim 21, wherein each potential unit boundary defines two candidate units for each speech segment.

27. The machine-readable medium of claim 26, wherein a concatenation of the plurality of concatenations includes a candidate unit of a first segment linked to a candidate unit of a second segment.

28. The machine-readable medium of claim 26, wherein the plurality of concatenations includes all combinations of a first candidate unit of each segment with a second candidate unit of each segment.

29. The machine-readable medium of claim 21, wherein the plurality of speech segments includes speech segments which end in the middle of a first phoneme, and speech segments which begin in the middle of a first phoneme.

30. The machine-readable medium of claim 29, wherein the plurality of speech segments are stored in a voice table.

31. The machine-readable medium of claim 21, wherein the method further comprises:

recording speech input; and

identifying the speech segments within the speech input.

32. The machine-readable medium of claim 21, wherein the portions include centered pitch periods, the centered pitch periods derived from pitch periods of the segments.

33. The machine-readable medium of claim 32, wherein the feature vectors incorporate phase information of the portions.

34. The machine-readable medium of claim 33, wherein creating feature vectors comprises:

constructing a matrix W from the portions; and

decomposing the matrix W.

35. The machine-readable medium of claim 34, wherein the matrix W is a $(2(K-1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the potential unit boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments, U is the $(2(K-1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K-1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R << (2(K-1)+1)M$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

36. The machine-readable medium of claim 35, wherein the centered pitch periods are symmetrically zero padded to N samples.

37. The machine-readable medium of claim 35, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

38. The machine-readable medium of claim 37, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

39. The machine-readable medium of claim 38, wherein a discontinuity $d(S_1,S_2)$ between two candidate units, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a

centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

40. The machine-readable medium of claim 39, wherein the same closeness measure, *C*, is used for optimizing unit boundaries and for unit selection.

41. An apparatus comprising:

means for extracting portions from segment boundary regions of a plurality of speech segments, each segment boundary region based on a corresponding initial unit boundary;

means for creating feature vectors that represent the portions in a vector space;

for each of a plurality of potential unit boundaries within each segment boundary region, means for determining an average discontinuity based on distances between the feature vectors; and

for each segment, means for selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

42. The apparatus of claim 41, further comprising:

if all of the new unit boundaries are the same as the corresponding initial unit boundaries, means for setting the new unit boundaries as final unit boundaries for the segments.

43. The apparatus of claim 41, further comprising:

if any of the new unit boundaries are different from the corresponding initial unit boundaries, means for iteratively:

setting the new unit boundary as the initial unit boundary, and

performing the extracting, the creating, the determining and the selecting, until all of the new unit boundaries are the same as the corresponding initial unit boundaries.

44. The apparatus of claim 41, wherein the average discontinuity is determined over a plurality of concatenations.

45. The apparatus of claim 41, wherein the initial unit boundary is in the middle of a phoneme.

46. The apparatus of claim 41, wherein each potential unit boundary defines two candidate units for each speech segment.

47. The apparatus of claim 46, wherein a concatenation of the plurality of concatenations includes a candidate unit of a first segment linked to a candidate unit of a second segment.

48. The apparatus of claim 46, wherein the plurality of concatenations includes all combinations of a first candidate unit of each segment with a second candidate unit of each segment.

49. The apparatus of claim 41, wherein the plurality of speech segments includes speech segments which end in the middle of a first phoneme, and speech segments which begin in the middle of a first phoneme.

50. The apparatus of claim 49, wherein the plurality of speech segments are stored in a voice table.

51. The apparatus of claim 41, further comprising:

means for recording speech input; and

means for identifying the speech segments within the speech input.

52. The apparatus of claim 41, wherein the portions include centered pitch periods, the centered pitch periods derived from pitch periods of the segments.

53. The apparatus of claim 52, wherein the feature vectors incorporate phase information of the portions.

54. The apparatus of claim 53, wherein creating feature vectors comprises:

means for constructing a matrix W from the portions; and

means for decomposing the matrix W.

55. The apparatus of claim 54, wherein the matrix W is a $(2(K\text{-}1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the potential unit boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row vectors v_j $(1 \leq j \leq N)$, $R << (2(K\text{-}1)+1)M)$, and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

56. The apparatus of claim 55, wherein the centered pitch periods are symmetrically zero padded to N samples.

57. The apparatus of claim 55, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

58. The apparatus of claim 57, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

59. The apparatus of claim 58, wherein a discontinuity $d(S_1,S_2)$ between two candidate units, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

60. The apparatus of claim 59, wherein the same closeness measure, C, is used for optimizing unit boundaries and for unit selection.

61. A system comprising:

a processing unit coupled to a memory through a bus; and

a process executed from the memory by the processing unit to cause the processing unit to:

extract portions from segment boundary regions of a plurality of speech segments, each segment boundary region based on a corresponding initial unit boundary;

create feature vectors that represent the portions in a vector space;

for each of a plurality of potential unit boundaries within each segment boundary region, determine an average discontinuity based on distances between the feature vectors; and

for each segment, select the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

62. The system of claim 61, wherein the process further causes the processing unit to:

if all of the new unit boundaries are the same as the corresponding initial unit boundaries, set the new unit boundaries as final unit boundaries for the segments.

63. The system of claim 61, wherein the process further causes the processing unit to:

if any of the new unit boundaries are different from the corresponding initial unit boundaries, iteratively:

set the new unit boundary as the initial unit boundary, and

perform the extracting, the creating, the determining and the selecting,

until all of the new unit boundaries are the same as the corresponding initial unit boundaries.

64. The system of claim 61, wherein the average discontinuity is determined over a plurality of concatenations.

65. The system of claim 61, wherein the initial unit boundary is in the middle of a phoneme.

66. The system of claim 61, wherein each potential unit boundary defines two candidate units for each speech segment.

67. The system of claim 66, wherein a concatenation of the plurality of concatenations includes a candidate unit of a first segment linked to a candidate unit of a second segment.

68. The system of claim 66, wherein the plurality of concatenations includes all combinations of a first candidate unit of each segment with a second candidate unit of each segment.

69. The system of claim 61, wherein the plurality of speech segments includes speech segments which end in the middle of a first phoneme, and speech segments which begin in the middle of a first phoneme.

70. The system of claim 69, wherein the plurality of speech segments are stored in a voice table.

71. The system of claim 61, wherein the process further causes the processing unit to:

record speech input; and

identify the speech segments within the speech input.

72. The system of claim 61, wherein the portions include centered pitch periods, the centered pitch periods derived from pitch periods of the segments.

73. The system of claim 72, wherein the feature vectors incorporate phase information of the portions.

74. The system of claim 73, wherein the process further causes the processing unit, when creating feature vectors, to:

construct a matrix W from the portions; and

decompose the matrix W.

75. The system of claim 74, wherein the matrix W is a $(2(K\text{-}1)+1)M \times N$ matrix represented by

$$W = U \Sigma V^T$$

where K-1 is the number of centered pitch periods near the potential unit boundary extracted from each segment, N is the maximum number of samples among the centered pitch periods, M is the number of segments, U is the $(2(K\text{-}1)+1)M \times R$ left singular matrix with row vectors u_i $(1 \leq i \leq (2(K\text{-}1)+1)M)$, Σ is the $R \times R$ diagonal matrix of singular values $s_1 \geq s_2 \geq \ldots \geq s_R > 0$, V is the $N \times R$ right singular matrix with row

vectors v_j ($1 \leq j \leq N$), $R << (2(K\text{-}1)+1)M$), and ^{T} denotes matrix transposition, wherein decomposing the matrix W comprises performing a singular value decomposition of W.

76. The system of claim 75, wherein the centered pitch periods are symmetrically zero padded to N samples.

77. The system of claim 75, wherein a feature vector $\bar{u}_i$ is calculated as

$$\bar{u}_i = u_i \Sigma$$

where u_i is a row vector associated with a centered pitch period i, and Σ is the singular diagonal matrix.

78. The system of claim 77, wherein the distance between two feature vectors is determined by a metric comprising a closeness measure, C, between two feature vectors, $\bar{u}_k$ and $\bar{u}_l$, wherein C is calculated as

$$C(\bar{u}_k, \bar{u}_l) = \cos(u_k\Sigma, u_l\Sigma) = \frac{u_k \Sigma^2 u_l^T}{\| u_k \Sigma \| \| u_l \Sigma \|}$$

for any $1 \leq k, l \leq (2(K\text{-}1)+1)M$.

79. The system of claim 78, wherein a discontinuity $d(S_1,S_2)$ between two candidate units, S_1 and S_2, is calculated as

$$d(S_1,S_2) = C(u_{\pi_{-1}}, u_{\delta_0}) + C(u_{\delta_0}, u_{\sigma_1}) - C(u_{\pi_{-1}}, u_{\pi_0}) - C(u_{\sigma_0}, u_{\sigma_1})$$

where $u_{\pi_{-1}}$ is a feature vector associated with a centered pitch period π_{-1}, u_{δ_0} is a feature vector associated with a centered pitch period δ_0, u_{σ_1} is a feature vector

associated with a centered pitch period σ_1, u_{π_0} is a feature vector associated with a centered pitch period π_0, and u_{σ_0} is a feature vector associated with a centered pitch period σ_0.

80. The system of claim 79, wherein the same closeness measure, *C*, is used for optimizing unit boundaries and for unit selection.

81. A machine-implemented method comprising:

setting an initial unit boundary for each segment of a plurality of speech segments, each initial unit boundary defining a segment boundary region and a plurality of potential unit boundaries within each segment boundary region;

for each segment, determining an average discontinuity over a plurality of concatenations of candidate units defined by the potential unit boundaries;

for each segment, selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

82. The machine-implemented method of claim 81, further comprising iteratively performing:

for each segment, setting the new unit boundary as the initial unit boundary; and

performing the determining and the selecting,

until all of the new unit boundaries for each segment are the same as the corresponding initial unit boundaries for each segment.

83. The machine-implemented method of claim 82, wherein determining the average discontinuity comprises:

constructing a matrix from time-domain samples of segment boundary regions; and

decomposing the matrix.

84. The machine-implemented method of claim 83, wherein the time-domain samples include centered pitch periods.

85. A machine-readable medium having instructions to cause a machine to perform a machine-implemented method comprising:

setting an initial unit boundary for each segment of a plurality of speech segments, each initial unit boundary defining a segment boundary region and a plurality of potential unit boundaries within each segment boundary region;

for each segment, determining an average discontinuity over a plurality of concatenations of candidate units defined by the potential unit boundaries;

for each segment, selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

86. The machine-readable medium of claim 85, the method further comprising iteratively performing:

for each segment, setting the new unit boundary as the initial unit boundary; and

performing the determining and the selecting,

until all of the new unit boundaries for each segment are the same as the corresponding initial unit boundaries for each segment.

87. The machine-readable medium of claim 86, wherein determining the average discontinuity comprises:

constructing a matrix from time-domain samples of segment boundary regions; and

decomposing the matrix.

88. The machine-readable medium of claim 87, wherein the time-domain samples include centered pitch periods.

89. An apparatus comprising:

means for setting an initial unit boundary for each segment of a plurality of speech segments, each initial unit boundary defining a segment boundary region and a plurality of potential unit boundaries within each segment boundary region;

for each segment, means for determining an average discontinuity over a plurality of concatenations of candidate units defined by the potential unit boundaries;

for each segment, means for selecting the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

90. The apparatus of claim 89, further comprising means for iteratively performing:

for each segment, means for setting the new unit boundary as the initial unit boundary; and

means for performing the determining and the selecting,

until all of the new unit boundaries for each segment are the same as the corresponding initial unit boundaries for each segment.

91. The apparatus of claim 90, wherein determining the average discontinuity comprises:

means for constructing a matrix from time-domain samples of segment boundary regions; and

means for decomposing the matrix.

92. The apparatus of claim 91, wherein the time-domain samples include centered pitch periods.

93. A system comprising:

a processing unit coupled to a memory through a bus; and

a process executed from the memory by the processing unit to cause the processing unit to:

set an initial unit boundary for each segment of a plurality of speech segments, each initial unit boundary defining a segment boundary region and a plurality of potential unit boundaries within each segment boundary region;

for each segment, determine an average discontinuity over a plurality of concatenations of candidate units defined by the potential unit boundaries;

for each segment, select the potential unit boundary associated with a minimum average discontinuity as a new unit boundary.

94. The system of claim 93, wherein the process further causes the processing unit to iteratively:

for each segment, set the new unit boundary as the initial unit boundary; and

perform the determining and the selecting,

until all of the new unit boundaries for each segment are the same as the corresponding initial unit boundaries for each segment.

95. The system of claim 94, wherein the process further causes the processing unit, when determining the average discontinuity, to:

construct a matrix from time-domain samples of segment boundary regions; and

decompose the matrix.

96. The system of claim 95, wherein the time-domain samples include centered pitch periods.